UNIVERSITY OF BATH

Citation for published version:
Simoes, B, Conceição, N, Kelsh, R & Cancela, ML 2014, 'Identification of cis-regulatory elements in the upstream regions of zebrafish runx3 gene through an in silico analysis: Implications for function. ', *Journal of Applied Ichthyology*, vol. 30, no. 4, pp. 661-670. https://doi.org/10.1111/jai.12536

DOI:
10.1111/jai.12536

Publication date:
2014

Document Version
Peer reviewed version

Link to publication

This is the peer reviewed version of the following article: Simoes, B, Conceição, N, Kelsh, R & Cancela, ML 2014, 'Identification of cis-regulatory elements in the upstream regions of zebrafish runx3 gene through an in silico analysis: Implications for function. ' Journal of Applied Ichthyology, vol 30, no. 4, pp. 661-670., which has been published in final form at http://dx.doi.org/10.1111/jai.12536. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

Received:16 January 2014
Accepted: 03 April 2014
Editor: Jean-Yves SIRE

Original article

Identification of *cis*-regulatory elements in the upstream regions of zebrafish *runx3* through *in silico* analysis: Implications for function

Running title: *in silico* analysis of *runx3* promoter regions

By B. Simões[1, 2, 3 #], N. Conceição[3 #], R.N. Kelsh[4] and M.L. Cancela[1, 3]

[1]Department of Biomedical Sciences and Medicine and [2]PhD program in Biomedical Sciences, University of Algarve, Faro, Portugal; [3]Centre of Marine Sciences/CCMAR/CIMAR-LA, University of Algarve, Faro, Portugal; [4]Department of Biology and Biochemistry, University of Bath, Claverton Down, United Kingdom

[#]Both authors contributed equally

Summary

RUNX3 encodes a member of the runt domain family of transcription factors. In mammals this family includes three genes (*RUNX1-3*) and their protein products function as context-dependent transcription factors, either transcriptional activators or repressors, during developmental processes such as hematopoiesis, neurogenesis, and osteogenesis; all are proto-oncogenes or tumour suppressors. All three genes were shown to be transcribed from two promoters, giving rise to protein products bearing either the P1 or the P2 N-termini, translated respectively from transcripts originating from the distal (P1)- or the proximal (P2)-promoters. Understanding their differential regulation and interaction may help explain how RUNX factors contribute to such different and often opposing biological processes. In this study, we have identified putative molecular players affecting zebrafish *runx3* transcription by using a computational approach to search for cis-regulatory transcription factor binding sites (TFBSs) in the *runx3* promoter regions of zebrafish (*Danio rerio*) and fugu (*Takifugu rubripes*). From the data obtained it was possible to identify the sites most likely involved in regulating expression of *runx3* in zebrafish. Our comparative approach reduced substantially the number of putative TFBSs in the *runx3* promoter regions; reassuringly, published TFs identified as transcriptional regulators of *Runx3* are confirmed by our *in silico* analysis. Our data now provides the basis for focused *in vitro* and/or *in vivo* experimental tests of the transcriptional regulatory activities of strong candidate regulators of zebrafish *runx3*.

Keywords: transcription factor, runx3, transfection, promoter regions, comparative analysis, transcription regulation

Introduction

RUNX3 encodes a member of the runt domain family of transcription factors, which also include *RUNX1* and *RUNX2*. RUNX proteins can bind DNA as a monomer to the core sequence 5'-PyGPyGGT-3' found in a number of enhancers and promoters, but their affinity for DNA is enhanced when the RUNX protein forms a heterodimer with its non-DNA binding partner CBFβ (Ogawa et al., 1993; Bae et al., 1994). The RUNX proteins also interact with other transcription factors, thus modulating their activity. Despite the recognized importance of this family in gene transcription, little is known about the factors regulating *RUNX3* transcription. Like the other two *RUNX* genes, *RUNX3* was shown to be transcribed from two promoters (Ghozi et al., 1996; Xiao et al., 1998; Rini and Calabi, 2001), giving rise to RUNX3 protein products bearing either the P1 or the P2 N-termini, resulting from transcripts derived from the distal (P1) or the proximal (P2) promoters, respectively (Bangsow et al., 2001; Rini and Calabi, 2001). The identification of several RUNX binding sites in the *RUNX* promoter regions (Ghozi et al., 1996; Levanon et al., 2001; Bangsow et al., 2001) led to the demonstration that auto- and cross-regulation of *RUNX* expression by RUNX proteins was likely to contribute to their regulation (Drissi et al., 2000; Spender et al., 2005). However, little work has addressed the regulatory processes that determine when RUNX proteins bind to the promoters of the genes of the other two family members to inhibit their expression in a kind of intrafamilial competition, nor when each RUNX protein acts mainly on its own promoter either promoting or inhibiting its own transcription, for example, to stabilise its levels of expression. Spender et al. (2005) have shown that in human B lymphoid cell lines, RUNX3 represses *RUNX1* expression, thus contributing to their mutually exclusive expression in those cells. In this case, RUNX3 represses the *RUNX1* P1 promoter by binding specifically to the conserved RUNX sites located near

74 the transcription start site of that promoter, thus confirming that cross-regulation
75 between different RUNX family members is a means of controlling *RUNX* expression
76 (Spender et al., 2005). The demonstration that RUNX transcription factors can be
77 regulated by other members of the RUNX family may help explain their diverse
78 functions and has important implications for the interpretation of pathologies associated
79 with *RUNX* gene knockout or amplification. RUNX family proteins can function as
80 context-dependent transcription factors during diverse developmental processes such as
81 hematopoiesis (de Bruijn and Speck, 2004), neurogenesis (Li et al., 2002; Fainaru et al.,
82 2004), and osteogenesis (Karsenty, 2000; Komori, 2005). RUNX2 and RUNX3 have
83 also been shown to regulate chondrocyte differentiation and maturation (Yoshida et al.,
84 2004). In zebrafish, loss of function of *runx3* was shown to lead to severe reduction of
85 head cartilage at 4 days post-fertilization (dpf) (Flores et al., 2006; Dalcq et al., 2012).
86 Furthermore, it was shown that a regulatory cascade formed by Runx3-Egr1-Sox9b
87 controls late chondrogenesis by reducing expression of Follistatin A, a BMP inhibitor
88 (Dalcq et al., 2012). This down-regulation allows the correct activation of BMP
89 signalling required for expression of *runx2b* in developing chondrocytes (Dalcq et al.,
90 2012). These observations were further investigated by Larbuisson et al. (2013). Using
91 loss of function studies these authors observed cartilage defects in Fgfr1a or Fgfr2
92 morphants that could be rescued by expression of exogenous Runx3 or Egr1. Recently,
93 using RNA-sequencing of Atlantic salmon notochord during segmentation it was also
94 shown that *runx3* was one of the genes expressed during - and implicated in - tissue
95 mineralisation, alongside other genes such as the chondroblast-specific *sox6*, *sox5* and
96 *sox9* (Wang et al., 2014)

97 Recently, we have cloned the full-length cDNA sequence of the zebrafish *runx3*,
98 observed the tissue distribution pattern and analyzed their bioinformatic features

(Simões, unpubl. results). With the aim to characterize the genomic structure and to analyze the promoter activities, we have cloned the 5′-flanking regions of the *runx3* (P1 and P2). In this study we have analysed the promoter activities of the 5′-flanking regions of the zebrafish *runx3* (P1 and P2) and identified putative transcriptional regulators of zebrafish *runx3* by using a computational approach to search for *cis*-regulatory transcription factor binding sites (TFBSs) in the promoter regions (P1 and P2) of the *runx3* gene from zebrafish (*Danio rerio*) and fugu (*Takifugu rubripes*). Our *in silico* strategy provides a quick way to identify the most promising candidates among the large number of TFs that might potentially regulate zebrafish *runx3 in vivo*. Testing the functionality of these sites *in vitro* and *in vivo* will then be the priority for future studies.

Materials and Methods

Cell culture, transient transfection and luciferase assay

C6 cells (rat glioma cell line) were maintained in F-12K Nutrient Mixture Medium supplemented with 2.5% fetal bovine serum, 15% horse serum and 1% penicillin/streptomycin. U2OS cells (human osteosarcoma cell line) were maintained in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum, 2mM L-Glutamine and 1% penicillin/streptomycin. For both cell lines, incubation was carried out at 37 °C in a humidified atmosphere containing 5% CO_2. The media, FBS, antibiotics and glutamine were obtained from Invitrogen. For all transfections, the C6 and the U2OS cells were seeded into 24-well plates at the density of 5 x 10^4 cells/well and 3 × 10^4 cells/well, respectively. Following 16 h of incubation, when the cells were about 70-80% confluent, transient transfection of the plasmids was carried out using Lipofectamine® LTX with Plus™ Reagent (Invitrogen) for C6 cells

and X-tremeGENE HP DNA Transfection Reagent (Roche) for U2OS cells. To normalize the transfection efficiency, the pRL-null vector (Promega) encoding Renilla luciferase was co-transfected at the ratio of 1:10 relative to pGL3-basic vector. After incubation for 48 h, cells were harvested. The luciferase activities were measured by the Dual-luciferase Reporter Assay System (Promega). The data were normalized by calculating the ratio of the specific activity of firefly luciferase to that of Renilla luciferase.

Sequence collection

Sequence databases at GenBank (www.ncbi.nlm.nih.gov) and Ensembl (release v72; www.ensembl.org) were searched for annotated Runx3 sequences derived from zebrafish and fugu. The promoter sequences from these two genes were extracted for analysis by selecting 5000 base pairs (bp) upstream of the known translation initiation site (TIS) giving rise to both isoforms, P1 and P2. This length of sequence provided a reasonable assurance of containing the target gene's TFBSs. The promoter sequences were masked for repetitive elements by the program RepeatMasker (www.repeatmasker.org) with the default mode.

Comparative analysis of promoter TFBSs

For each zebrafish-fugu orthologous promoter pair, the DNA Block Aligner (DBA) software (www.ebi.ac.uk/Tools/psa/promoterwise/) was used to extract blocks of aligned sequence using the default parameter settings based on the postulation that conserved regulatory blocks may be regions important for regulation of the gene. DBA alignments between orthologous promoters vary substantially, in many cases having no significant alignment, but others having several large sections of aligning sequences. This also clearly shows that many genes have multiple aligning blocks, sometimes spaced quite widely apart in the 5 kb region. With this software it is possible to obtain

four different types of conserved blocks of a certain degree of similarity: type A, 60-70%; type B, 70-80%; type C, 80%-90%; and type D, 90-100%.

The promoter sequences (P1 and P2) of the zebrafish *runx3* gene were then assessed for TFBSs by running MatInspector (http://www.genomatix.de/) against TF binding site position weight matrices (PWM). For this study we used the default settings for the core similarity 0.75 and for matrix similarity 0.80.

For the multiple alignment plus prediction of TFBSs in the set of identified conserved blocks present within the promoters, we used the DiAlignTF software (http://www.genomatix.de/). We retained the same settings as used for the TFBSs prediction in the promoters (score similarity 0.75 and matrix similarity 0.8) and all the common TFBS matches located in aligned regions were determined. Then we used MatInspector for quantification of TFBSs common to all input sequences. The percentage of retention of putative TFBSs was calculated comparing the number of a given TFBS in the promoter to that from the conserved blocks in both zebrafish and fugu promoters.

Results

Promoter activity analysis and prediction of transcription factor binding sites in zebrafish runx3

To investigate the activity of the promoter regions (P1 and P2) of zebrafish *runx3*, we performed luciferase assays following the transient transfection of the human osteosarcoma U2OS cells and rat C6 glioma cells (**Fig. 1**). The U2OS cell line was shown to express low levels of RUNX3 (Lai and Mager, 2012) and it was previously reported that RUNX3 expression is significantly decreased in human glioma (Mei et al., 2011). We have used P1 (from -5094 to -17 of the TIS starting as MASN) and P2 (from

-3930 to -474 of the TIS starting as MHIPV) reporter constructs generated by inserting PCR fragments into the pGL3-basic vector (Simões, unpubl. results). Both the constructs exhibited higher luciferase activities than negative control of the pGL3-basic vector. The P2 promoter construct showed about 36-fold and 26-fold higher luciferase activities than pGL3-basic in U2OS and C6 cells respectively, while the P1 promoter construct showed about 7-fold and 5-fold higher activity than the empty vector in U2OS and C6 cells, respectively. Therefore, we expect that differences in the sequence of each promoter could affect their ability to function as a promoter.

In order to verify the relationship between sequence variation and promoter activity, we analyzed TFBSs in the zebrafish *runx3* promoter regions by using TRANSFAC® Public 6.0. We observed a huge number of putative TFBSs in both promoters, some of which identified in both P1 and P2, while others specifically found either in P1 or P2 promoter regions. To identify from this list of TFBSs, which ones are more likely to be involved in the regulation of each promoter we decided to perform a comparative promoter analysis.

Comparison of zebrafish and fugu runx3 promoter regions

To identify the likely regulatory regions, we analyzed the conservation of the promoter regions of the zebrafish and fugu *runx3* genes using the DBA (DNA Block Aligner) web server. The output of DBA not only identifies discrete conserved blocks but also classifies them into four levels of conservation (A-D category, A showing lowest (60-70%) and D showing highest (90-100%) conservation). Among the 5 kb promoter sequences upstream of the TIS, an average of ten per cent of the total length of the P1 promoter region from zebrafish and fugu could be aligned by DBA; 0.7%, 3.2%, 1.9%, and 3.9% being of the A-D category, respectively. For the P2 promoter region, only 3.3% was aligned by DBA; 0.5%, 0.9% and 1.9% being of the B-D categories,

199 respectively. We plotted the sequence conservation as a function of the distance from
200 the zebrafish *runx3* translation start sites (**Fig. 2).** Sequence conservation in the first
201 1000 bp was distinctly higher (**Fig. 2C**), a finding fully consistent with the typical
202 pattern for protein coding genes (Conceição et al., 2009). Furthermore, in addition to
203 being more abundant, conserved blocks in the first 1000 bp tended to be more
204 conserved (i.e. Type D, with 90-100% sequence identity) in both promoters (**Fig. 2**).
205 Interestingly, we observed that the P1 promoter region shows more conserved blocks
206 (23 blocks) than the P2 promoter (eight blocks) (**Table 1**), suggesting that the P1
207 promoter sequence is more conserved compared with P2 and that there are more
208 conserved *cis*-acting regulatory elements in the P1 than in the P2 promoter region.
209 ***Analysis of regulatory elements using MatInspector***
210 To analyze the conservation of the *cis* elements between the zebrafish and fugu 5 kb
211 *runx3* promoter regions, the zebrafish promoter sequences and the set of identified
212 conserved blocks present within the zebrafish and fugu promoters were compared for
213 TFBSs. Our comparison reveals the conservation of multiple potential *cis* elements
214 between fugu and zebrafish promoters. Although TFBSs are abundant in all sequences
215 assessed, relatively few show conservation between zebrafish and fugu. Thus, for
216 zebrafish, this criterion reduces the 3478 TFBSs in the P1 promoter sequence to only
217 142 in the conserved blocks, and the 8544 TFBSs in the P2 promoter sequence to only
218 104 in the conserved blocks. Thus, by identifying and analysing only conserved blocks,
219 an average of 96% and 99% of all TFBSs in, respectively, the P1 and P2 promoter
220 regions were eliminated (**Fig. 3**). Likewise, this approach also substantially reduces the
221 number of TF families implicated in *runx3* regulation, (from 170 to 66 (a reduction of
222 61 %) and from 175 to 49 (a reduction of 72 %), in P1 and P2 promoters, respectively).

To determine whether our comparative screening of TFBSs was likely to have identified TFs (**Table 1**) with roles in *runx3* regulation, we searched our lists for the presence of TFs previously characterized as regulating *RUNX3*. Key roles have been shown for signal transducer and activator of transcription STAT (Park et al., 2010), Sp1 transcription factor /TEA domain family member 2 (Sp1/ETF) (Bangsow et al., 2001), T helper transcription factor (Th-POK) (Egawa et al., 2009), CBF1/Suppressor of Hairless/Lag1 (CSL) (Fu et al., 2011), interferon regulatory factor 4 (IRF4) (Cao et al., 2010), E-twenty-six (Ets1) (Zamisch et al., 2009), cAMP-response element-binding protein (Creb) (Lim et al., 2011), microphthalmia-associated transcription factor (Mitf) (Hoek et al., 2008), Twist subfamily of class B bHLH transcription factors (Scl/Tal1) (Landry et al., 2008), Brn POU domain factor (Brn3a) (Dykes et al., 2010; 2011), recombination signal binding protein-J kappa (Rbpj) (Fu et al., 2011) and runt-related transcription factors (Runx) (Drissi et al., 2000) in regulation of mammalian *RUNX3.* Of these eight TFs, Ets1, Stat, Creb, Mitf, Scl, Brn3a, Rbpj and Runx show conserved binding sites in the conserved sequence blocks in the *runx3* promoter sequences analysed here, consistent with them having a functional role in the regulation of *runx3* in fish. These TFs are then prime candidates for future functional studies assessing their ability to bind to and regulate activity of *runx3 in vivo*.

Discussion

In the present study we provide evidence for the transcriptional activity of zebrafish *runx3* promoters in two different cell lines, using *in vitro* transient transfection experiments. These findings support an earlier report from our laboratory (Simões, unpubl. results). To gain insight into the regulatory mechanism of the *runx3* gene, the sequences of the genomic fragments (named P1 and P2) were analyzed *in silico* for

potential recognition sites to transcription factors. Our analysis identified numerous putative *cis*-regulatory elements that may serve as targets for sequence-specific enhancer/silencer transcription factors.

We have then used the DBA algorithm to obtain comparative alignment between zebrafish and fugu *runx3* promoter regions in order to detect conserved sequence blocks and then used MatInspector to determine putative TFBSs in those blocks, so as to enrich for likely functionally relevant TFBSs. Our *in silico* analysis of zebrafish *runx3* P1 and P2 promoter regions provides important clues as to factors likely to be involved in regulation of *runx3* expression. Although MatInspector can find most true positive TFBS matches in a promoter region (Cartharius et al., 2005), not all sites found are necessarily functional in a particular biological context. A first step in examining functionality is a comparative promoter analysis. The alignment obtained with DBA was then assessed for conserved TFBSs by DiAlignTF, a combination of MatInspector with the multiple alignment program DiAlign (Morgenstern *et al*., 1998). DiAlignTF displays TFBSs located at the same position within the alignment and then it can be used to reduce the list of potential TFBSs to the most likely functional matches. From the data obtained it was possible to identify the sites most likely involved in regulating expression of *runx3* in zebrafish. While a number of pathways regulating RUNX activity have been delineated, transcription factors binding to *RUNX* promoters are only beginning to be identified. From the list of 86 putative TFBSs families retained after the comparative analysis, CREB (family V$CREB) (Lim et al., 2011), Mitf (family V$MITF) (Hoek et al., 2008), Brn3a (family V$BRNF) (Dykes et al., 2010; 2011), Rbp-j (family V$RBPF) (Fu et al., 2011), Scl/Tal1 (family V$HAND) (Landry et al., 2008), Ets1 (family V$ETSF) (Zamisch et al., 2009), Stat (family V$STAT) (Park et al., 2010) and Runx (family V$HAML) (Drissi et al., 2000; Spender et al., 2005) are

described in the literature as regulating *RUNX3*. In addition to these data indicating likely conservation of a regulatory function for these TFs between mammals and fish, our *in silico* analysis identified a number of novel potential regulatory TFs for the zebrafish *runx3* promoters (Fig. 3). As described in Materials and Methods, several TFs were found to be retained within the conserved blocks (between zebrafish and fugu promoters) analysed. By focusing only on those, we selected a set having a retained score of 15 % or higher (see Fig. 3), which included Brn-5 POU domain factors (family V$BRN5), Cas interacting zinc finger (family V$CIZF), Runx or Human acute myelogenous leukemia factors (family V$HAML), PAX-4/PAX-6 paired domain binding sites (family V$PAX6), Spalt-like transcription factor 2 (family V$SAL2) and Sterol regulatory element binding proteins (family V$SREB) for P1 promoter and TFIIB or RNA polymerase II transcription factor II B (family O$TF2B) for P2 promoter. Interestingly, available data links some of these with either skeletal or neuronal development. In the context of the P1 promoter, five of these TFs have a function in neurogenesis and two in skeletogenesis. Brn-5 POU domain factors (Brn-5) is expressed in many central nervous system (CNS) neuron populations and may function as a transcriptional regulator involved in specifying the mature phenotype of CNS neurons (Cui and Bulleit, 1998). Spalt-like transcription factor 2 (Sall2) also plays a role in neuronal development (Pincheira et al., 2009) and is the only member of the family suggested to act as a tumor suppressor (Li et al., 2001; Ma et al., 2001). Sterol regulatory element binding protein 2 (Srebp2) was shown to interact with the *acetoacetyl-CoA synthetase* (*AACS*) promoter and knockdown experiments showed that SREBP-2 regulates *AACS* expression during neurite outgrowth in the neuroblastoma Neuro-2a cell line (Hasegawa et al., 2012). Paired box 6 (Pax6) also is expressed during neurogenesis (Gan et al., 2013), and it functions as a transcription factor with a major

298 role in eye and brain development from Drosophila to humans (Callaerts et al., 1999;
299 van Heyningen and Williamson, 2002). Lleras-Forero et al. (2013) showed CNS *Pax6b*
300 expression in zebrafish. The human acute myelogenous leukemia factors (Haml; also
301 known as a runt-related Runx/AML protein) function as context-dependent transcription
302 factors during developmental processes such as hematopoiesis, neurogenesis, and
303 osteogenesis (Westendorf and Hiebert, 1999). Runx expression was shown in subtypes
304 of dorsal root ganglion (DRG) neurons, suggesting their involvement in lamina-specific
305 afferent differentiation and maturation (Inoue et al., 2003) and Runx2 and Runx3 have
306 also been shown to regulate chondrocyte differentiation and maturation (Yoshida et al.,
307 2004; Komori, 2005). In zebrafish, we have shown that Runx2 was able to transactivate
308 the promoter of *osteocalcin*, an osteoblastic marker gene (Pinto et al., 2005), as well as
309 the promoter of *collagen Xα1*, a chondrocyte marker gene (Simões et al., 2006).
310 Zebrafish *runx3* expression was observed in neuronal tissues including the trigeminal
311 ganglia and Rohon-Beard neurons (Kalev-Zylinska et al., 2003) and also in the
312 craniofacial region (Flores et al., 2006). Cas-interacting zinc finger protein (CIZ) is one
313 of the suppressors of BMP signalling in osteoblastic differentiation (Shen et al., 2002).
314 Besides these last two highly conserved TFBSs, our data show the occurrence in the
315 conserved sequences of binding sites for many more TFs that are described as having a
316 role in skeletogenesis. These factors include NF-kB (Wu et al., 2011), NF-YB (Chen et
317 al., 2009), NFATc1 (Lambertini et al., 2008), Ets-1 (Wenke et al., 2006), and Sox5 and
318 Sox9 (Yang et al., 2011). Curiously, all these TFBSs are only detected in the P1
319 promoter and not in the P2 promoter.

320 Of the two putative TFBSs more conserved in the P2 promoter, that for the odd-skipped
321 related (Osr) zinc finger transcription factor is notable since it was suggested to be
322 involved in bone formation (Kawai et al., 2007). The other relates to TFIIB, a

component of the basal transcription complex. Several other TFs identified in our analysis for the P2 promoter are known to play critical roles in zebrafish development, e.g. *dlx* (distal-less homeodomain; family V$DLXF) genes play a key role in the patterning of the forebrain, in peripheral structures of the head, and in the fins (Akimenko et al., 1994); *mef2* (myocyte enhancer factor 2; family V$MEF2) genes are essential for heart development (Hinits et al., 2012) and in cranial neural crest for proper head skeletal patterning (Miller et al., 2007); CREB (cAMP response element-binding protein; family V$CREB) have a role in neural development (Dworkin et al., 2007); Nkx6 (NK6 homeobox; family V$NKX6) proteins specify one zebrafish primary motoneuron subtype (Hutchinson et al., 2007); Pax3 (paired box 3; family V$PAX3) is induced early during neural development in progenitors of the dorsal spinal cord (Moore et al., 2013), and Six3 (sine oculis homeobox homolog 3; family V$SIX3) are involved in the left-right brain patterning (Inbal et al., 2007).

Since it is known that all Runx protein family members bind to the same DNA core sequence, their temporal and/or spatial expression has to be tightly regulated. Other investigators have reported that the two promoter regions, P1 and P2, regulate *Runx3* expression in a cell type-specific manner (Bangsow et al., 2001; Egawa, 2009). In mice Soung et al. (2007) showed that both Runx3 isoforms are expressed and regulated during chondrocyte differentiation, while Yoshida et al. (2004) showed that *Runx3* mRNA was detected in both CD8+ and CD4+ T cells, but only the CD8+ population expressed the P1 transcript isoform and detectable levels of RUNX3.

Comparison of the zebrafish *runx2b* proximal promoter sequence that we identified and cloned previously (Pinto et al., 2005) with both those of zebrafish *runx3* showed some common consensus binding motifs, namely for NFAT, CREB, RUNX, and CBF1. They are thus possible candidates for regulating expression of *runx3*. Of interest are the two

putative RUNX-binding sites present in the P1 regulatory region of all three *RUNX* genes, at the beginning of the 5'UTR, perfectly conserved in mammals. These RUNX-binding sites were previously shown by independent studies to have an effect on the transcriptional regulation of *RUNX* genes, either positively or negatively, through the binding of RUNX proteins (Levanon et al, 1998; Ducy et al, 1999; Drissi et al, 2000; Bangsow et al, 2001; Levanon and Groner, 2004; Spender et al, 2005). In the present analysis we also found two Runx-binding sites in the *runx3* P1 promoter that are conserved between zebrafish and fugu. Taken together this may indicate important regulatory roles such as cross-regulation and/or auto-regulation.

In conclusion, our comparative *in silico* analysis of zebrafish *runx3* gene promoter regions, using the DBA and DiAlignTF softwares, predicts strong candidates TFBSs likely to contribute to regulation of *runx3* transcription. These TFBSs include binding sites for TFs already known from work in mammals as transcriptional regulators of *Runx3*, but also include novel TFs. Thus, our data likely provide a powerful tool to guide future dissection of *runx3* transcriptional regulation *in vitro* and/or *in vivo*.

Acknowledgments

NC and BS are supported, respectively, by post-doctoral and PhD grants from FCT (SFRH/BPD/48206/2008 and SFRH/BD/38083/2007). This research was partially supported by the European Regional Development Fund (ERDF) through the COMPETE - Operational Competitiveness Program and national funds through FCT – Foundation for Science and Technology, under the project "PEst-C/MAR/LA0015/2011.

372 **References:**

373 Akimenko, M. A.; Ekker, M.; Wegner, J.; Lin, W.; Westerfield, M., 1994:
374 Combinatorial expression of three zebrafish genes related to distal-less: part of a
375 homeoboxgene code for the head. J. Neurosci. **14**, 3475-3486.

376 Bae, S. C.; Ogawa, E.; Maruyama, M.; Oka, H.; Satake, M.; Shigesada, K.; Jenkins, N.
377 A.; Gilbert, D. J.; Copeland, N. G.; Ito, Y., 1994: PEBP2 alpha B/mouse AML1
378 consists of multiple isoforms that possess differential transactivation potentials.
379 Mol. Cell Biol. **14**, 3242-3252.

380 Bangsow, C.; Rubins, N.; Glusman G.; Bernstein, Y.; Negreanu, V.; Goldenberg, D.;
381 Lotem, J.; Ben-Asher, E.; Lancet, D.; Levanon, D.; Groner, Y., 2001: The
382 RUNX3 gene - sequence, structure and regulated expression. Gene **279**, 221-232.

383 Callaerts, P.; Munoz-Marmol ,A. M.; Glardon, S.; Castillo, E.; Sun, H.; Li, W. H.;
384 Gehring, W. J.; Salo, E., 1999: Isolation and expression of a pax-6 gene in the
385 regenerating and intact planarian Dugesia(G)tigrina. Proc. Natl. Acad. Sci. USA
386 **96**, 558-563.

387 Cao, Y.; Li, H.; Sun, Y.; Chen, X.; Liu, H.; Gao, X.; Liu, X., 2010: Interferon
388 regulatory factor 4 regulates thymocyte differentiation by repressing Runx3
389 expression. Eur. J. Immunol. **40**, 3198-3209.

390 Cartharius, K.; Frech, K.; Grote, K.; Klocke, B.; Haltmeier, M.; Klingenhoff, A.; Frisch,
391 M.; Bayerlein M.; Werner, T., 2005: MatInspector and beyond: promoter analysis
392 based on transcription factor binding sites. Bioinformatics **21**, 2933–2942.

393 Chen, Y. H.; Lin, Y. T.; Lee, G. H., 2009: Novel and unexpected functions of zebrafish
394 CCAAT box binding transcription factor (NF-Y) B subunit during cartilages
395 development. Bone **44**, 777-784.

396 Conceição, N.; Cox, C. J.; Simões, B.; Viegas, M.; Cancela, M. L., 2009: Comparative
397 promoter analysis and its application to the identification of candidate regulatory
398 factors of cartilage-expressed genes. J. Appl. Ichthyol. **26**, 245-250.

399 Cui, H.; Bulleit, R. F., 1998: Expression of the POU transcription factor Brn-5 is an
400 early event in the terminal differentiation of CNS neurons. J. Neurosci. Res. **52**,
401 625-632.

402 Dalcq, J.; Pasque, V.; Ghaye, A.; Larbuisson, A.; Motte, P.; Martial, J. A.; Muller, M.,
403 2012: RUNX3, EGR1 and SOX9B form a regulatory cascade required to

404 modulate BMP-signaling during cranial cartilage development in zebrafish. PLoS
405 One **7**, e50140.
406 De Bruijn, M. F.; Speck, N. A., 2004: Core-binding factors in hematopoiesis and
407 immune function. Oncogene **23**, 4238-4248.
408 Drissi, H.; Luc, Q.; Shakoori, R.; Lopes, S. C. S.; Choi, J. Y.; Terry, A.; Hu, M.; Jones,
409 S.; Neil, J. C.; Lian, J. B.; Stein, J. L.; Van Wijnen, A. J.; Stein, G. S., 2000:
410 Transcriptional autoregulation of the bone related CBFA1/RUNX2 gene. J Cell
411 Physiol. **184**, 341-350
412 Ducy, P.; Starbuck, M.; Priemel, M.; Shen, J.; Pinero, G.; Geoffroy, V.; Amling, M.;
413 Karsenty, G., 1999: A Cbfa1-dependent genetic pathway controls bone formation
414 beyond embryonic development. Genes Dev. **13**, 1025-1036.
415 Dworkin, S.; Heath, J. K.; deJong-Curtain, T. A.; Hogan, B. M.; Lieschke, G. J.;
416 Malaterre, J.; Ramsay, R. G.; Mantamadiotis, T., 2007: CREB activity modulates
417 neural cell proliferation, midbrain-hindbrain organization and patterning in
418 zebrafish. Dev Biol. **307**, 127-141.
419 Dykes, I. M.; Lanier, J.; Eng, S. R.; Turner, E. E., 2010: Brn3a regulates neuronal
420 subtype specification in the trigeminal ganglion by promoting Runx expression
421 during sensory differentiation. Neural Dev. **5**, 1-18.
422 Dykes, I. M.; Tempest, L.; Lee, S. I.; Turner, E. E., 2011: Brn3a and Islet1 act
423 epistatically to regulate the gene expression program of sensory differentiation. J.
424 Neurosci. **31**, 9789-9799.
425 Egawa, T., 2009: Runx and ThPOK: a balancing act to regulate thymocyte lineage
426 commitment. J. Cell Biochem. **107**, 1037-1045.
427 Fainaru, O.; Woolf, E.; Lotem, J.; Yarmus, M.; Brenner, O.; Goldenberg, D.; Negreanu,
428 V.; Bernstein, Y.; Levanon, D.; Jung, S.; Groner, Y., 2004: Runx3 regulates
429 mouse TGF-beta-mediated dendritic cell function and its absence results in airway
430 inflammation. EMBO J. **23**, 969-979.
431 Flores, M. V.; Lam, E. Y.; Crosier, P.; Crosier, K., 2006: A hierarchy of Runx
432 transcription factors modulate the onset of chondrogenesis in craniofacial
433 endochondral bones in zebrafish. Dev. Dyn. **235**, 3166-3176.
434 Fu, Y.; Chang, A. C.; Fournier, M.; Chang, L.; Niessen, K.; Karsan, A., 2011: RUNX3
435 maintains the mesenchymal phenotype after termination of the Notch signal. J.
436 Biol. Chem. **286**, 11803-11813.

437 Gan , Q.; Lee, A.; Suzuki, R.; Yamagami, T.; Stokes, A.; Nguyen, B. C.; Pleasure, D.;
438 Wang, J.; Chen, H. W.; Zhou, C. J., 2013: Pax6 mediates ß-catenin signaling for
439 self-renewal and neurogenesis by neocortical radial glial stem cells. Stem Cells
440 doi: 10.1002/stem.1561.
441 Ghozi, M. C.; Bernstein, Y.; Negreanu, V.; Levanon, D.; Groner, Y., 1996: Expression
442 of the human acute myeloid leukemia gene AML1 is regulated by two promoter
443 regions. Proc. Natl. Acad. Sci. USA **93**, 1935-1940.
444 Hasegawa, S.; Kume, H.; Iinuma, S.; Yamasaki, M.; Takahashi, N.; Fukui, T., 2012:
445 Acetoacetyl-CoA synthetase is essential for normal neuronal development.
446 Biochem. Biophys. Res. Commun. **427**, 398-403.
447 Hinits, Y.; Pan, L.; Walker, C.; Dowd, J.; Moens, C. B.; Hughes, S. M., 2012: Zebrafish
448 Mef2ca and Mef2cb are essential for both first and second heart
449 fieldcardiomyocyte differentiation. Dev. Biol. **369**, 199-210.
450 Hutchinson, S. A.; Cheesman, S. E.; Hale, L. A.; Boone, J. Q.; Eisen, J. S., 2007: Nkx6
451 proteins specify one zebrafish primary motoneuron subtype by regulating late
452 islet1 expression. Development **134**, 1671-1677.
453 Inbal, A.; Kim, S. H.; Shin, J.; Solnica-Krezel, L., 2007: Six3 represses nodal activity to
454 establish early brain asymmetry in zebrafish. Neuron **55**, 407-415.
455 Inoue, K.; Ozaki, S.; Ito, K.; Iseda, T.; Kawaguchi, S.; Ogawa, M.; Bae, S. C.;
456 Yamashita, N.; Itohara, S.; Kudo, N.; Ito, Y., 2003: Runx3 is essential for the
457 target-specific axon pathfinding of trkc-expressing dorsal root ganglion neurons.
458 Blood Cells Mol. Dis. **30,** 157-160.
459 Kalev-Zylinska, M. L.; Horsfield, J. A.; Flores, M. V.; Postlethwait, J. H.; Chau, J. Y.;
460 Cattin, P. M.; Vitas, M. R.; Crosier, P. S.; Crosier, K. E. 2003: Runx3 is required
461 for hematopoietic development in zebrafish. Dev. Dyn. **228**, 323-336.
462 Karsenty, G., 2000: Role of Cbfa1 in osteoblast differentiation and function. Semin.
463 Cell Dev. Biol. **11**, 343–346.
464 Kawai, S.; Yamauchi, M.; Wakisaka, S.; Ooshima, T.; Amano, A., 2007: Zinc-finger
465 transcription factor odd-skipped related 2 is one of the regulators in osteoblast
466 proliferation and bone formation. J. Bone Miner. Res. **22**, 1362-1372.
467 Komori, T., 2005: Regulation of Skeletal Development by the Runx Family of
468 Transcription Factors. J. Cell Biochem. **95**, 445–453.

Lai, C. B.; Mager, D. L., 2012: Role of runt-related transcription factor 3 (RUNX3) in transcription regulation of natural cytotoxicity receptor 1 (NCR1/NKp46), an activating natural killer (NK) cell receptor. J. Biol. Chem. **287**, 7324-7334.

Lambertini, E.; Penolazzi, L.; Tavanti, E.; Pocaterra, B.; Schincaglia, G. P.; Torreggiani, E.; Franceschetti, T.; Vecchiattini, R.; Gambari, R.; Piva, R., 2008: Modulation of expression of specific transcription factors involved in the bone microenvironment. Minerva Biotec. **20**, 69-77.

Landry, J. R.; Kinston, S.; Knezevic, K.; de Bruijn, M. F.; Wilson, N.; Nottingham, W. T.; Peitz, M.; Edenhofer, F.; Pimanda, J. E.; Ottersbach, K.; Göttgens, B., 2008: Runx genes are direct targets of Scl/Tal1 in the yolk sac and fetal liver. Blood **111**, 3005-3014.

Larbuisson, A.; Dalcq, J.; Martial, J. A.; Muller, M., 2013: Fgf receptors Fgfr1a and Fgfr2 control the function of pharyngeal endoderm in late cranial cartilage development. Differentiation S0301-4681(13)00063-7.

Levanon, D.; Goldstein, R. E.; Bernstein, Y.; Tang, H.; Goldenberg, D.; Stifani, S.; Paroush, Z.; Groner, Y., 1998: Transcriptional repression by AML1 and LEF-1 is mediated by the TLE/Groucho corepressors. Proc. Natl. Acad. Sci. USA. **95**, 11590-11595.

Levanon, D.; Brenner, O.; Negreanu, V.; Bettoun, D.; Woolf, E.; Eilam, R.; Lotem, J.; Gat, U.; Otto, F.; Speck, N.; Groner, Y., 2001: Spatial and temporal expression pattern of Runx3 (Am12) and Runx1 (Am11) indicates non-redundant functions during mouse embryogenesis. Mech. Dev. **109**, 413-417.

Levanon, D.; Groner, Y., 2004: Structure and regulated expression of mammalian RUNX genes. Oncogene **23**, 4211-4219.

Li, D.; Dower, K.; Ma, Y.; Tian, Y.; Benjamin, T. L., 2001: A tumor host range selection procedure identifies p150 (sal2) as a target of polyoma virus large T antigen. Proc. Natl. Acad. Sci. USA **98**, 14619-14624.

Li, Q. L.; Ito, K.; Sakakura, C.; Fukamachi, H.; Inoue, K.; Chi, X. Z.; Lee, K. Y.; Nomura, S.; Lee, C. W.; Han, S. B.; Kim, H. M.; Kim, W. J.; Yamamoto, H.; Yamashita, N.; Yano, T.; Ikeda, T.; Itohara, S.; Inazawa, J.; Abe, T.; Hagiwara, A.; Yamagishi, H.; Ooe, A.; Kaneda, A.; Sugimura, T.; Ushijima, T.; Bae, S. C.; Ito, Y., 2002: Causal relationship between the loss of RUNX3 expression and gastric cancer. Cell **109**, 113-124.

Lim, B.; Ju, H.; Kim, M.; Kang, C., 2011: Increased genetic susceptibility to intestinal-type gastric cancer is associated with increased activity of the RUNX3 distal promoter. Cancer **117**, 5161-5171.

Ma, Y.; Li, D.; Chai, L.; Luciani, A. M.; Ford, D.; Morgan, J.; Maizel, A. L., 2001: Cloning and characterization of two promoters for the human HSAL2 gene and their transcriptional repression by the Wilms tumor suppressor gene product. J. Biol. Chem. **276**, 48223-48230.

Mei, P. J.; Bai, J.; Liu, H.; Li, C.; Wu, Y. P.; Yu, Z. Q.; Zheng, J. N., 2011: RUNX3 expression is lost in glioma and its restoration causes drastic suppression of tumor invasion and migration. J. Cancer Res. Clin. Oncol. **137**, 1823-1830.

Miller, C. T.; Swartz, M. E.; Khuu, P. A.; Walker, M. B.; Eberhart, J. K.; Kimmel, C.B., 2007: mef2ca is required in cranial neural crest to effect Endothelin1 signaling in zebrafish. Dev. Biol. **308**, 144-157.

Moore, S.; Ribes, V.; Terriente, J.; Wilkinson, D.; Relaix, F.; Briscoe, J., 2013: Distinct regulatory mechanisms act to establish and maintain Pax3 expression in the developing neural tube. PLoS Genet. **9**, e1003811.

Morgenstern, B.; Frech, K.; Dress, A.; Werner, T., 1998: DIALIGN: finding local similarities by multiple sequence alignment. Bioinformatics **14**, 290–14294.

Park, J. H.; Adoro, S.; Guinter, T.; Erman, B.; Alag, A. S.; Catalfamo, M.; Kimura, M. Y.; Cui, Y.; Lucas, P. J.; Gress, R. E.; Kubo, M.; Hennighausen, L.; Feigenbaum, L.; Singer, A., 2010: Signaling by intrathymic cytokines, not T cell antigen receptors, specifies CD8 lineage choice and promotes the differentiation of cytotoxic-lineage T cells. Nat. Immunol. **11,** 257-264.

Pincheira, R.; Baerwald, M.; Dunbar, J. D.; Donner, D. B., 2009: Sall2 is a novel p75NTR-interacting protein that links NGF signalling to cell cycle progression and neurite outgrowth. Embo J. **28**, 261-273.

Pinto, J. P.; Conceição, N. M.; Viegas, C. S.; Leite, R. B.; Hurst, L. D.; Kelsh, R. N.; Cancela, M. L., 2005: Identification of a new pebp2alphaA2 isoform from zebrafish runx2 capable of inducing osteocalcin gene expression in vitro. J. Bone Miner. Res. **20**, 1440-1453.

Rini, D.; Calabi, F., 2001: Identification and comparative analysis of a second runx3 promoter. Gene **273**, 13-22.

Shen, Z. J.; Nakamoto, T.; Tsuji, K.; Nifuji, A.; Miyazono, K.; Komori, T.; Hirai, H.; Noda, M., 2002: Negative regulation of bone morphogenetic protein/Smad

signaling by Cas-interacting zinc finger protein in osteoblasts. J. Biol. Chem. **277**, 29840-29846.

Simões, B.; Conceição, N.; Viegas, C. S.; Pinto, J. P.; Gavaia, P. J.; Hurst, L. D.; Kelsh, R. N.; Cancela, M. L., 2006: Identification of a promoter element within the zebrafish colXalpha1 gene responsive to runx2 isoforms Osf2/Cbfa1 and til-1 but not to pebp2alphaA2. Calcif. Tissue Int. **79,** 230-244

Soung do, Y.; Dong, Y.; Wang, Y.; Zuscik, M. J.; Schwarz, E. M.; O'Keefe, R. J.; Drissi, H., 2007: Runx3/AML2/Cbfa3 regulates early and late chondrocyte differentiation. J Bone Miner Res. **22**, 1260-1270.

Spender, L. C.; Whiteman, H. J.; Karstegl, C. E.; Farrell, P.J., 2005: Transcriptional cross-regulation of RUNX1 by RUNX3 in human B cells. Oncogene **24**, 1873-1881.

van Heyningen, V.; Williamson, K. A., 2002: PAX6 in sensory development. Hum. Mol. Genet. **11**, 1161-1167.

Wang, S.; Wang, Q.; Crute, B. E.; Melnikova, I. N.; Keller, S. R.; Speck, N. A., 1993: Cloning and Characterization of Subunits of the T-Cell Receptor and Murine Leukemia Virus Enhancer Core-Binding Factor. Mol. Cell Biol. **13**, 3324-3339.

Wang, S.; Furmanek, T.; Kryvi, H.; Krossøy, C.; Totland, G. K.; Grotmol, S.; Wargelius, A., 2014: Transcriptome sequencing of Atlantic salmon (Salmo salar L.) notochord prior to development of the vertebrae provides clues to regulation of positional fate, chordoblast lineage and mineralisation. BMC Genomics **15**, 141.

Wenke, A. K.; Rothhammer, T.; Moser, M.; Bosserhoff, A. K., 2006: Regulation of integrin alpha10 expression in chondrocytes by the transcription factors AP-2epsilon and Ets-1. Biochem. Biophys. Res. Commun. **345**, 495-501.

Westendorf, J. J.; Hiebert, S. W., 1999: Mammalian Runt-Domain Proteins and Their Roles in Hematopoiesis, Osteogenesis, and Leukemia. J. Cell. Biochem. Supplements **32/33**, 51-58.

Wu, S.; Zang, W.; Li, X.; Sun, H., 2011: Proepithelin stimulates growth plate chondrogenesis via nuclear factor-{kappa}B-p65-dependent mechanisms. J. Biol. Chem. **286**, 24057-24067.

Xiao, Z. S.; Thomas, R.; Hinson, T. K.; Quarles, L. D., 1998: Genomic structure and isoform expression of the mouse, rat and human Cbfa1/Osf2 transcription factor. Gene **214**, 187-197.

570 Yang, H. N.; Park, J. S.; Woo, D. G.; Jeon, S. Y.; Do, H. J.; Lim, H. Y.; Kim, S. W.;
571 Kim, J. H.; Park, K. H., 2011: Chondrogenesis of mesenchymal stem cells and
572 dedifferentiated chondrocytes by transfection with SOX Trio genes. Biomaterials
573 **32**, 7695-7704.
574 Yoshida, C. A.; Yamamoto, H.; Fujita, T.; Furuichi, T.; Ito, K.; Inoue, K.; Yamana, K.;
575 Zanma, A.; Takada, K.; Ito, Y.; Komori, T., 2004: Runx2 and Runx3 are essential
576 for chondrocyte maturation, and Runx2 regulates limb growth through induction
577 of Indian hedgehog. Genes Dev. **18**, 952-963.
578 Zamisch, M.; Tian, L.; Grenningloh, R.; Xiong, Y.; Wildt, K. F.; Ehlers, M.; Ho, I. C.;
579 Bosselut, R., 2009: The transcription factor Ets1 is important for CD4 repression
580 and Runx3 up-regulation during CD8 T cell differentiation in the thymus. J. Exp.
581 Med. **206**, 2685-2699.
582
583
584 Author's address:
585 Natércia Conceição, CCMAR-University of Algarve, Campus de Gambelas, 8005-139
586 Faro, Portugal
587 E-mail: nconcei@ualg.pt
588
589

590 **Legends:**

591 Figure 1. Relative transcriptional activity of zebrafish *runx3* promoter constructs (P1
592 and P2) in U2OS and C6 cell lines. The results represent the ratio between firefly and
593 Renilla luciferase determined. The mean and the SD for at least three independent
594 transfections are shown. pGL3-Basic is the empty vector which lacks eukaryotic
595 promoter and enhancer sequences and served as a control.

596

597 Figure 2. Distribution of percentage of base-pairs located in block A, B, C, or D located
598 in zebrafish and fugu promoters, for each of five 1000 bp segments spanning up to -
599 5000 bp upstream of translation initiation site (TIS) in (a) P1, (b) P2 and (c) P1 and P2
600 *runx3* promoter regions.

601

602 Figure 3. Retention of putative TFBSs after comparative analysis. For each TF listed
603 along the *x*-axis, corresponding bars represent the percentage (*y*-axis) of putative TFBSs
604 originally identified by MatInspector that also survived after DiAlignTF comparative
605 analysis. The dashed line indicates the minimum percent chosen to consider TFBSs as
606 most frequent.

607

608 Figure 4. Representation of a DBA block obtained from the alignment of P1 (a) and P2
609 (a') *runx3* promoter regions between zebrafish (Dr) and fugu (Fr) and overview of
610 TFBS families detected by DiAlignTF on the conserved blocks analysed for P1 (b) and
611 P2 (b') promoters. (a, a') Examples of an alignment of one of the 23 blocks obtained for
612 P1 and 8 blocks obtained for P2 using the DBA software. The block position in the
613 respective promoter sequence is shown, considering the A of the translation initiation
614 codon as +1. The block type (type B and type C) is also represented as a bold letter next

615 to the consensus sequence identified between the two blocks. (b, b') Overview of the
616 TFBS conserved in the block showed as (a) or (a'), respectively for P1 or P2, detected
617 by MatInspector using DiAlignTF program. Only upper-case letters are considered to be
618 aligned. The colour code for each specific TFBS is shown above the alignment.
619
620 Table 1. Transcription factor families conserved in each block obtained from the
621 alignment of P1 and P2 *runx3* promoter regions between zebrafish (Dr) and fugu (Fr).
622
623 Supplementary Figure S1. TFBS families detected by DiAlignTF that are common in all
624 conserved blocks obtained from the alignment of *runx3* P1 (a) and P2 (b) promoter
625 regions between zebrafish and fugu. In each block is represented the alignment obtained
626 by DBA software (upper alignment) and the DiAlignTF output (lower alignment)
627 showing the TFBSs conserved in each block. The colour code for each specific TFBS is
628 shown above the alignment.
629